American Journal of Hematology 54:108–117 (1997)

Laboratory Diagnosis of Anemia and Related Diseases Using Multivariate Analysis

Shuichi Shiga,[1]* Toshihito Furukawa,[3] Itsuko Koyanagi,[1] Morihisa Yamagishi,[2] Yataro Yoshida,[2] Takayuki Takahashi,[4] Reiji Kannagi,[5] and Toru Mori[1]

[1]Central Clinical Laboratory, Kyoto University School of Medicine, Kyoto, Japan
[2]First Division of Internal Medicine, Kyoto University School of Medicine, Kyoto, Japan
[3]Department of Scientific Development Headquarters, SRL, Inc., Hachiogi, Tokyo, Japan
[4]Department of Medicine, Kobe General Hospital, Kobe, Japan
[5]Second Division of Pathology, Aichi Cancer Center, Nagoya, Japan

To establish a simple computer program for the laboratory diagnosis of anemia and related diseases, multivariate analyses were applied to the results of routine hematological laboratory tests obtained from 48 patients and 51 healthy volunteers. The patients studied were limited to those who had not been treated hematologically by the time of their first visit to our hospital, and their first data obtained in our laboratory were analyzed. Final diagnoses were aplastic anemia (AA) in 21, myelodysplastic syndrome (MDS) in 14, iron deficiency anemia (IDA) in 3, polycytemia vera (PV)in 3, and idiopathic thrombocytopenic purpura (ITP) in 7. Eight parameters, WBC, RBC, Hb, Ht, MCV, MCH, MCHC, and PLT, were transformed to normal distribution and then applied to principal component analysis to evaluate their independence. Very close relationships were observed between Ht and Hb, and between MCV and MCH. One each of these pairs was selected by discriminant analysis and two sets, RBC, MCH, Hb, PLT, and WBC, and RBC, MCV, Ht, PLT, and WBC, were obtained. Two canonical components gave good discrimination of these five diseases and also of normal subjects. When disease prediction was made using this analysis, 37 of 48 patients (77.1%) were predicted correctly, and furthermore, when two disease predictions were allowed, all patients were diagnosed properly. Some overlaps were observed in this two-dimensional coordinate system, especially of AA and MDS, and also with normal subjects. To improve the system further, the additional parameters of age and sex were added to construct a three-dimensional analysis which resulted in much clearer discrimination. The whole procedure described is being developed with subjects who are not taking medication. Subsequently, the general application of this analytical procedure should be limited to only those not on medications. In conclusion, this is in essence a demonstration project; however, this trial of laboratory diagnosis using routine hematological laboratory results appears to be promising. Further extension of the study by increasing numbers of patients and disorders studied, including secondary anemias, will allow the design of diagnostic software for use with personal computers at the sites of primary care. Am. J. Hematol. 54:108–117, 1997

Key words: computer diagnosis; multivariate analysis; anemia; hematology; laboratory diagnosis

INTRODUCTION

Recent progress in laboratory medicine has enabled accurate laboratory information to be obtained very rapidly. However, it may be difficult for a physician to interpret all laboratory information correctly, because too much and variable data are delivered every day. Therefore, to facilitate effective discrimination of patients at the sites of primary care, it would be desirable to have simple laboratory diagnosis systems utilizing routine laboratory results only. Such systems have been reported previously [1–4].

In the present study, we developed a computer program for the differential diagnosis of anemia and related disor-

Abbreviations: RBC, red blood cell count; Hb, hemoglobin concentration; Ht, hematocrit; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin content; MCHC, mean corpuscular hemoglobin concentration; PLT, platelet count; WBC, white blood cell count; AA, aplastic anemia; MDS, myelodysplastic syndrome; IDA, iron deficiency anemia; PV, polycytemia vera; ITP, idiopathic thrombocytopenic purpura.

*Correspondence to: Shuichi Shiga, M.T., Central Clinical Laboratory, Kyoto University Hospital, 54 Shogoin Kawaharacho, Sakyo-ku, Kyoto 606-01, Japan.

Received for publication 11 March 1996; Accepted 11 September 1996.

Contract grant sponsor: SRL, Inc.

ders. The relationship between two variables is usually expressed by a correlation coefficient. However, in the case of multiple variables it is rather difficult to demonstrate any similarity among variables, or to look for the underlying factors simply by calculating correlation coefficients. For this purpose, routine hematological laboratory results were analyzed using the discriminant analysis based on Wilks' lambda to devise an "expert" diagnostic algorithm. Such an analysis is not new but has been only relatively infrequently applied to hematological problems. In order to facilitate wide application, we used SAS, which is readily available to most laboratories. These methods provided the basis for an excellent and useful algorithm for the differentiation of various disorders of the hematopoietic system.

MATERIALS AND METHODS

Clinical Materials

Forty-eight patients suffering from anemia or related disorders were selected for the study. The patients were limited to those who had not been treated by the time of their first visit to our hospital. Since the results of hematological examinations can be altered easily by medications, it was critical to restrict the usage of data only to those obtained prior to any medication. The patients were 46.3 ± 19.2 (mean ± SD) years old, comprised of 25 males (51.5 ± 18.1 years old) and 23 females (40.6 ± 19.1 years old). Their diagnoses were established through various clinical findings and laboratory results, and included 21 cases of aplastic anemia (AA), 14 of myelodysplastic syndrome (MDS), 3 of iron deficiency anemia (IDA), 3 of polycytemia vera (PV),and 7 of idiopathic thrombocytopenic purpura (ITP). Routine hematological examination results at the time of the first visit were analyzed (Table I).

Control Subjects

As the control subjects, 51 healthy adult employees in our hospital volunteered to participate in this study and donated blood samples. The control subjects were 33.7 ± 13.1 years old and were comprised of 18 males (36.5 ± 13.8 years old) and 33 females (32.1 ± 12.6 years old). Their test results, shown as the highest value, lowest value, and mean and SD of eight parameters, respectively, were as follows: WBC, 8.3, 3.4, and 5.5 ± 1.2 ($\times 1^3/\mu l$); RBC, 5.51, 3.75, and 4.56 ± 0.39 ($\times 10^6/\mu l$); Hb, 16.8, 11.7, and 13.9 ± 1.3 (g/dl); Ht, 51.2, 35.7, and 41.4 ± 3.8 (%); MCV, 106.2, 84.2, and 90.9 ± 4.2 (fl); MCH, 35.6, 28.1, and 30.4 ± 1.5 (pg); MCHC, 34.3, 32.4, and 33.4 ± 0.5 (%); and PLT, 420, 165, and 247 ± 51 ($\times 10^3/\mu l$).

These subjects were much younger than the patient groups. Age and sex were the major prognostic factors on the hematological laboratory results. After extensive analyses of these, we are now using fractional reference values routinely for each of our hematological laboratory parameters [17]. In the second analysis, these data were also adjusted for normalization by age and sex.

Statistical Analysis

Some analyses were performed after normalizing transformations, because some of the laboratory result distributions were skewed. Accordingly, WBC was log-transformed, PLT was log-log transformed, and MCV and MCH were transformed by their squares.

Relationships among these hematological laboratory results were analyzed using principal component analysis [5–7] (PRINCOMP procedure using SAS). The purpose of a principal component analysis is to derive a small number of linear combinations (principal components) from a set of variables that retain as much of the information in the original variables as possible. A principal component is a coordinate obtained after condensing the information of each variable (Eigen value) maximally. Plotting of individual values on a principal component system will reflect the cumulative proportion of these variables. The first principal component (PRIN1) has the largest variance (largest information) of any unit-length linear combination of the observed variables. The jth principal component (PRINj) has the largest variance of any unit-length linear combination orthogonal to the first j-1 principal component.

The combinations of laboratory results useful for the diagnosis of the five disorders were selected by discriminant analysis [8–11] based on Wilks' lambda (SAS; STEPDISC procedure). Discriminant analysis is a useful statistical method, analyzing multiple laboratory results of one individual to find out a particular group among several groups to which the individual belongs. Discriminant function was used to classify the observations into two or more known groups on the basis of test values. Classification was performed by the parametric method, using a measure of generalized squared distance (Mahalanobis' generalized distance). This distance can derive the probability distance between the center of the group and specific observation of other groups. Wilks' lambda is the classification statistic based on a linear regression between variable and variance of classification which shows the relationship of within-group and between-group variance. After selection of laboratory results, statistical models for the prediction of five diseases were achieved by canonical analysis (SAS; CANDISC procedure [12–16]). Canonical discriminant analysis derives a linear combination of the variable that has the maximal multiple correlation with the groups. This maximal multiple correlation is called the first canonical correlation (CAN1). The variable defined by the linear combination is the first canonical component. The second canonical correlation (CAN2) is obtained by finding the linear combina-

TABLE I. Hematological Laboratory Data From 48 Patients

Case	Sex	Age	Diagnosis	WBC	RBC	Hb	Ht	MCV	MCH	MCHC	PLT
1	F	53	AA	3.7	3.79	12.2	37.6	99.2	32.2	32.4	26
2	F	49	AA	2.2	2.00	6.8	19.8	99.0	34.0	34.3	17
3	F	15	AA	1.9	1.70	5.6	16.9	99.7	33.2	33.2	17
4	F	40	AA	2.6	2.46	7.7	24.2	98.4	31.3	31.8	52
5	M	32	AA	1.8	1.71	7.0	20.0	117.1	41.2	35.2	33
6	F	61	AA	4.8	2.98	10.4	30.5	102.3	34.9	34.1	13
7	M	39	AA	2.7	0.87	3.7	11.0	126.4	42.5	33.5	12
8	M	29	AA	2.5	1.45	6.2	18.7	129.0	42.8	33.2	14
9	M	52	AA	4.7	3.21	12.7	38.3	119.3	39.6	33.2	25
10	F	24	AA	5.7	3.12	10.7	31.8	101.9	34.3	33.7	108
11	F	33	AA	3.0	2.98	11.3	33.0	110.8	37.9	34.2	41
12	F	20	AA	2.9	2.92	9.1	26.7	91.4	31.2	34.1	54
13	F	60	AA	4.4	1.66	6.6	18.7	112.6	39.9	35.4	19
14	F	17	AA	2.4	2.28	7.1	21.6	94.7	31.1	32.9	10
15	M	63	AA	1.6	2.51	8.7	24.9	99.2	34.7	34.9	38
16	F	40	AA	1.7	3.10	10.4	29.5	95.2	33.5	35.3	18
17	F	14	AA	1.8	2.02	7.5	22.5	111.4	37.1	33.3	26
18	M	56	AA	1.3	0.87	3.7	10.3	118.4	42.5	35.9	9
19	M	22	AA	1.8	2.41	8.4	25.1	104.1	34.9	33.5	23
20	M	21	AA	3.5	2.42	9.6	29.0	119.8	39.7	33.1	6
21	M	61	AA	3.1	3.35	13.1	37.9	113.1	39.1	34.6	60
22	M	74	MDS	2.6	2.05	7.0	20.7	100.9	34.1	33.8	204
23	F	65	MDS	1.6	3.09	10.6	32.0	103.7	34.4	33.2	116
24	M	71	MDS	2.2	1.62	6.6	19.0	117.3	40.7	34.7	70
25	F	72	MDS	3.0	2.42	8.2	24.7	102.1	33.9	33.2	114
26	M	50	MDS	2.2	2.62	8.5	25.5	97.5	32.4	33.3	12
27	M	40	MDS	5.3	3.98	14.8	43.7	109.8	37.2	33.9	492
28	M	19	MDS	2.0	1.77	7.0	20.7	116.4	39.2	33.7	21
29	M	63	MDS	3.1	2.25	8.1	24.1	106.9	35.7	33.4	44
30	F	39	MDS	3.1	2.38	7.1	22.9	96.2	30.0	31.0	18
31	F	78	MDS	2.4	2.25	6.6	20.3	90.2	29.4	32.6	253
32	M	83	MDS	1.0	3.76	10.2	31.0	82.3	27.0	32.8	65
33	M	71	MDS	1.5	2.10	6.9	19.7	93.8	33.0	35.2	30
34	M	64	MDS	5.3	4.58	14.7	44.3	96.7	32.1	33.2	128
35	M	70	MDS	0.5	1.54	5.6	16.8	108.6	36.3	33.4	52
36	F	23	IDA	3.7	4.37	8.2	27.3	62.5	18.8	30.0	315
37	F	47	IDA	5.2	3.93	10.3	31.4	79.9	26.2	32.8	321
38	F	41	IDA	5.0	3.80	8.7	27.2	71.6	22.9	32.0	229
39	M	58	PV	22.7	6.40	13.8	43.8	68.4	21.6	31.5	1,400
40	M	33	PV	8.1	5.88	17.1	50.4	85.7	29.1	33.9	88
41	M	62	PV	18.9	5.21	11.6	37.0	71.0	22.3	31.4	1,254
42	M	59	ITP	6.5	4.32	14.5	43.2	100.0	33.6	33.6	105
43	M	55	ITP	3.6	4.68	14.6	43.7	93.2	31.2	33.5	28
44	M	40	ITP	10.0	5.33	16.3	48.1	90.2	30.7	34.0	20
45	F	18	ITP	13.5	4.37	13.5	39.9	91.3	30.9	33.8	44
46	F	24	ITP	4.3	4.67	13.8	41.6	89.1	29.6	33.2	76
47	F	48	ITP	4.0	3.80	11.7	34.7	91.2	30.8	33.8	126
48	F	53	ITP	7.0	4.65	14.5	41.8	89.8	31.2	34.7	14

tion uncorrelated with CAN1. The process of extracting canonical variables can be repeated until the number of canonical components equals the number of original laboratory results or the number of groups minus one.

At first, we constructed models using laboratory results only, and then adjustments for prognostic factors (age and sex [17]) were applied for further improvement.

All statistical analyses were performed using SAS programs.

RESULTS

Similarity Among Laboratory Results

According to the results of principal component analysis, we defined two major factors which would explain independent clinical parameters properly: factors indicating hematopoietic potential, and erythrocyte indices. These two factors were found to explain approximately 80% of all information available. Figure 1 shows distribu-

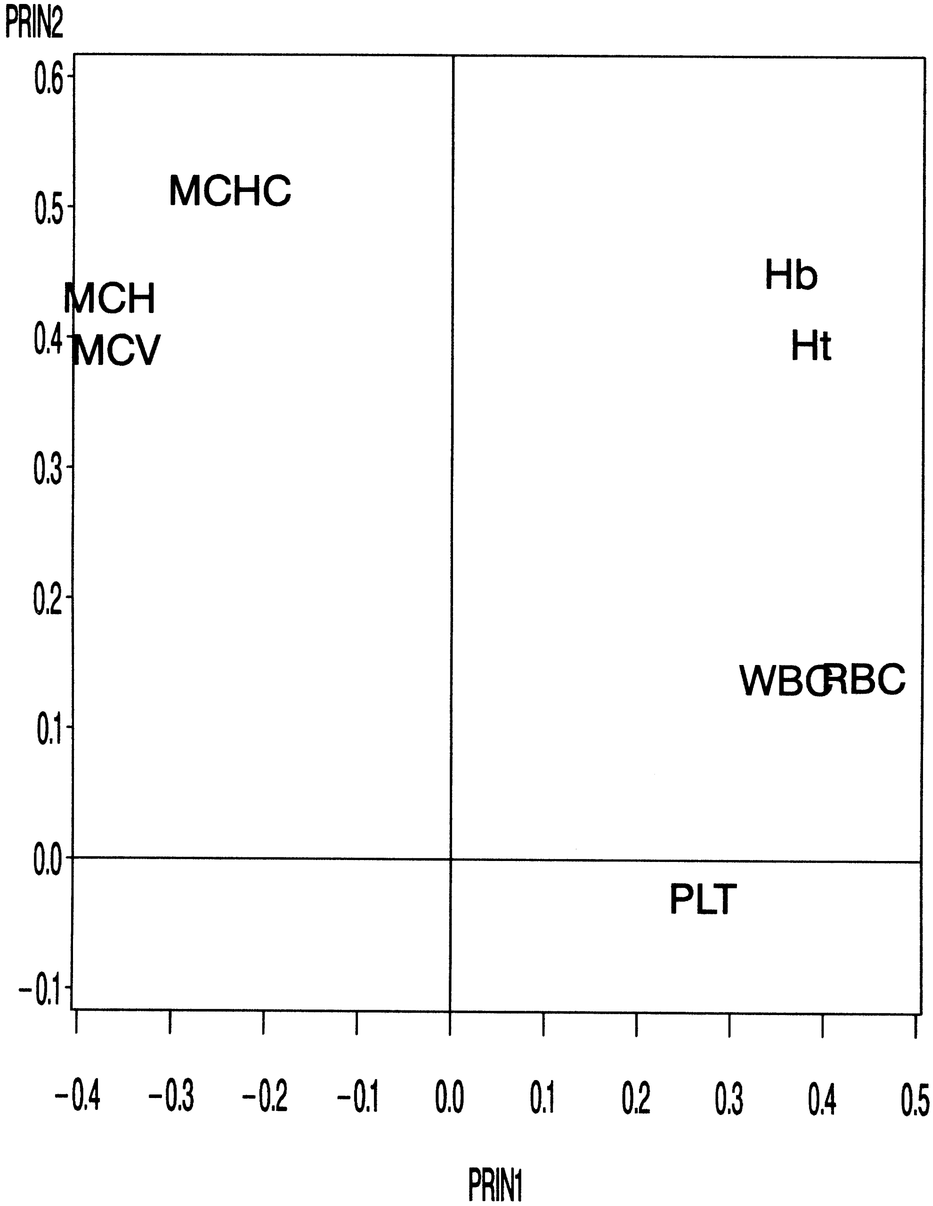

Fig. 1. Relationships between tests: principal component analysis. PRIN1 shows first principal component, and PRIN2 shows second. Plot of the test shows the Eigen vector, which is the correlation coefficient between each principal component and the test variables. This figure shows the distance on the map to demonstrate the relationship among the tests. Tests located nearby indicate that findings resembled each other on this analysis.

tion patterns of these eight parameters, plotting principal component 1 (PRIN1) vs. PRIN2 on a coordinate system. MCH and MCV gave very similar information, as did Hb and Ht. Therefore, one each of these pairs appeared to be sufficient for further analysis.

Selection of Laboratory Results Useful for Disease Prediction

As described in Materials and Methods, two sets of laboratory results useful for diagnosing five diseases were selected by discriminant analysis. The first picked up a set of four results, RBC, MCH, Hb, and PLT, and the second put more weight on another set comprised of RBC, MCV, Ht, and PLT. WBC is considered to give different useful information in other disorders, and we added WBC to each of the above sets.

In classification based on the generalized square distance, the overall ratio of correct prediction for the five disorders was 77.1% (37 of 48 cases), using either of the above sets (Table I). In other words, the false prediction

TABLE II. Disease Prediction in Model 1 Using RBC, MCH, Hb, PLT, and WBC

Case	Diagnosis	AA	IDA	ITP	MDS	PV	Prediction
1	AA	0.27	0.00	0.60	0.13	0.00	ITP
2	AA	0.68	0.00	0.00	0.32	0.00	AA
3	AA	0.61	0.00	0.00	0.39	0.00	AA
4	AA	0.25	0.00	0.00	0.75	0.00	MDS
5	AA	0.64	0.00	0.00	0.36	0.00	AA
6	AA	0.84	0.00	0.09	0.07	0.00	AA
7	AA	0.95	0.00	0.00	0.05	0.00	AA
8	AA	0.94	0.00	0.00	0.06	0.00	AA
9	AA	0.84	0.00	0.06	0.11	0.00	AA
10	AA	0.35	0.00	0.06	0.59	0.00	MDS
11	AA	0.61	0.00	0.01	0.38	0.00	AA
12	AA	0.27	0.00	0.02	0.71	0.00	MDS
13	AA	0.91	0.00	0.00	0.09	0.00	AA
14	AA	0.83	0.00	0.01	0.16	0.00	AA
15	AA	0.32	0.00	0.00	0.68	0.00	MDS
16	AA	0.60	0.00	0.02	0.38	0.00	AA
17	AA	0.57	0.00	0.00	0.43	0.00	AA
18	AA	0.94	0.00	0.00	0.06	0.00	AA
19	AA	0.55	0.00	0.00	0.45	0.00	AA
20	AA	0.98	0.00	0.01	0.01	0.00	AA
21	AA	0.54	0.00	0.02	0.43	0.00	AA
22	MDS	0.09	0.00	0.00	0.91	0.00	MDS
23	MDS	0.11	0.00	0.00	0.89	0.00	MDS
24	MDS	0.42	0.00	0.00	0.58	0.00	MDS
25	MDS	0.18	0.00	0.00	0.82	0.00	MDS
26	MDS	0.78	0.00	0.01	0.20	0.00	AA
27	MDS	0.16	0.00	0.11	0.73	0.00	MDS
28	MDS	0.75	0.00	0.00	0.25	0.00	AA
29	MDS	0.48	0.00	0.00	0.51	0.00	MDS
30	MDS	0.64	0.00	0.01	0.35	0.00	AA
31	MDS	0.04	0.00	0.00	0.96	0.00	MDS
32	MDS	0.05	0.04	0.08	0.83	0.00	MDS
33	MDS	0.31	0.00	0.00	0.69	0.00	MDS
34	MDS	0.02	0.00	0.94	0.04	0.00	ITP
35	MDS	0.08	0.00	0.00	0.92	0.00	MDS
36	IDA	0.00	0.94	0.00	0.00	0.06	IDA
37	IDA	0.00	0.94	0.05	0.01	0.00	IDA
38	IDA	0.00	1.00	0.00	0.00	0.00	IDA
39	PV	0.00	0.00	0.00	0.00	1.00	PV
40	PV	0.00	0.02	0.95	0.00	0.04	ITP
41	PV	0.00	0.00	0.00	0.00	1.00	PV
42	ITP	0.06	0.00	0.87	0.07	0.00	ITP
43	ITP	0.01	0.00	0.98	0.01	0.00	ITP
44	ITP	0.00	0.00	1.00	0.00	0.00	ITP
45	ITP	0.00	0.00	0.99	0.00	0.00	ITP
46	ITP	0.00	0.01	0.98	0.01	0.00	ITP
47	ITP	0.11	0.01	0.40	0.49	0.00	MDS
48	ITP	0.00	0.00	1.00	0.00	0.00	ITP

rate was 22.9%. Table II shows predicted probability of the five disorders in these patients using discriminant functions. Among 21 patients with AA, 16 (76.1%) were predicted correctly but 5 were mistakenly classified as MDS (n = 4) or ITP (n = 1). The rather high false prediction rate of MDS (19.0%) indicates a partial common property between these two disorders. Extensive prediction was defined as the classification in which one patient was predicted to have one of two disorders with the largest or the next largest posterior probability. Thus, from the viewpoint of extensive prediction, all 21 patients were diagnosed properly. In 14 patients with MDS, 10 were diagnosed correctly and 4 were indicated to have AA (n = 3) or ITP (n = 1). Again, extensive prediction gave correct results, and false prediction of AA was high as 21.4%. Two of 3 PV patients were diagnosed correctly, and 1 was predicted to have ITP. All but 1 of 7 patients with ITP were diagnosed correctly, 1 was indicated to have MDS with a 49% probability, and extensive prediction was good. All 3 patients with IDA were predicted

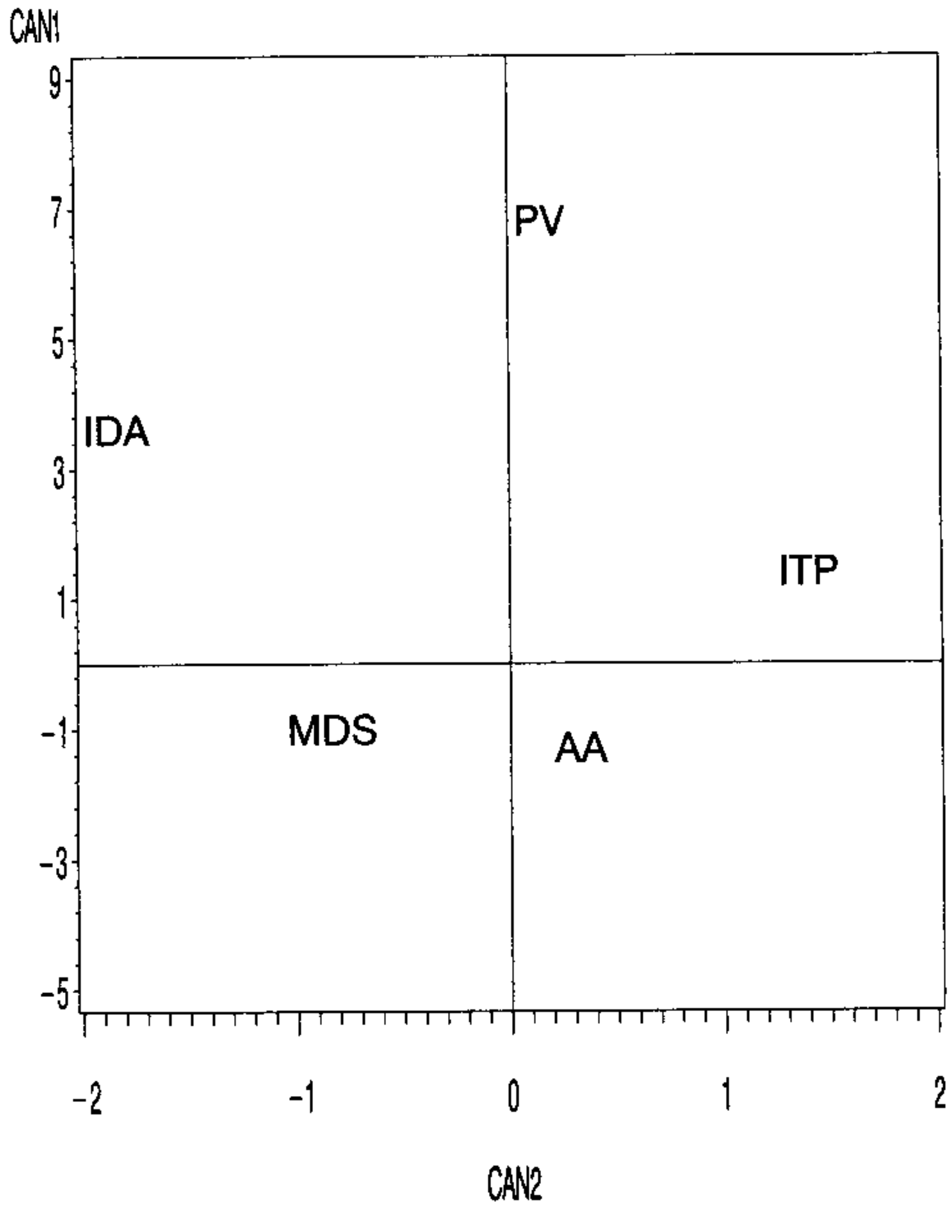

Fig. 2. Distribution of mean canonical disease scores in model 2 using RBC, MCH, Hb, PLT, and WBL. Plots show the canonical score of the average of test values obtained by five diseases. Canonical score is the linear combination of the test value. The coefficient of the linear combination was calculated by canonical analysis. This plot shows the relationship of the diseases based on the test.

correctly. If extensive prediction is allowed, all 48 patients were diagnosed properly.

Results of Canonical Analysis

Using a two-dimensional canonical analysis, mean weights of the five disorders were shown to be distant from each other, as shown in Figure 2. However, disease prediction in individual patients was not fully correct, indicating possible overlap among distribution areas of each disorder. As shown in Figure 3, IDA, PV, and ITP appeared to be distributed almost independently from each other, but MDS and AA overlapped significantly, especially in the middle portion.

Distribution of Control Subjects and Three-Dimensional Analysis

When data from 51 control subjects were added and analyzed together, most of the calculated values for the control subjects were distributed close to the center of the coordinate, and differed markedly from those of the

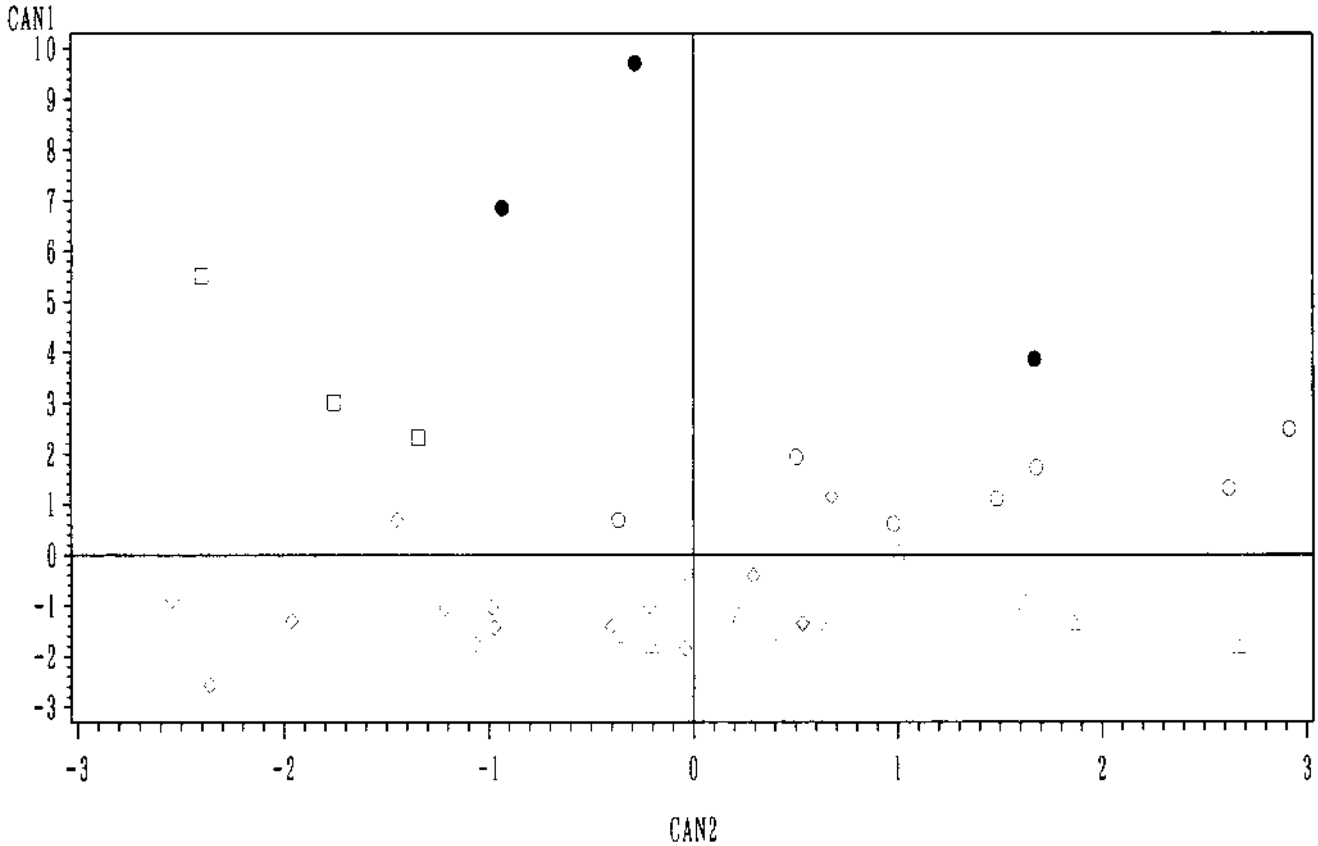

Fig. 3. Distribution of canonical scores of patients with model 2. Plots show the canonical scores of individual patient data as defined by the diseases. Triangle, AA; Square, IDA; open circle, ITP; diamond, MDS; closed circle, PV.

patients (Fig. 4). However, there were certain overlaps with data from diseased patients. To improve discrimination further, adjustments for prognostic factors (age and sex) were added, and disease prediction was then assessed by three-dimensional canonical analysis. As shown in Figure 5, control subjects were discriminated much more clearly from patients, and further differentiation between diseases, including AA and MDS, appeared much better than disease discrimination shown in Figures 3 or 4.

DISCUSSION

Multivariate analyses have often been applied to establish disease prediction systems using laboratory results only [8,18]. There have been several trials of computer-aided diagnosis of anemia and related disorders using laboratory databases [1–4]. A simple expert system has been applied, especially to microcytic anemia, giving a limited success for discrimination [3,4]. It is much more desirable to develop a widely applicable discrimination system, useful at the site of primary care. To establish an effective system for disease prediction, careful selection of the database which is used during development is critical with regard to: 1) selection of patients, and 2) selection of laboratory results. Previously, we attempted to analyze hematological laboratory results obtained from more than 250 patients with various disorders of the hematopoietic system. These results, however, were almost impossible to analyze, because alterations of data after medication greatly disturbed further analysis in detail. In this study, we restricted the data to those obtained prior to medication. Therefore, we could not analyze as many patients, although this selection allowed analysis in greater detail. On the other hand, useful laboratory results should be defined as those which contribute significantly and independently to the disorder(s). Multiple combinations of laboratory results with similar clinical significance do not improve diagnostic accuracy.

Through the principal component analyses among eight hematological parameters, two pairs of laboratory results, Hb with Ht and MCH with MCV, were shown to be almost equipollent. RBC and PLT showed significance individually. RBC and WBC were also located very close to each other, but these cannot be taken as equipollent. WBC showed no particular significance in this series of five-disorder analysis; however, WBC is considered

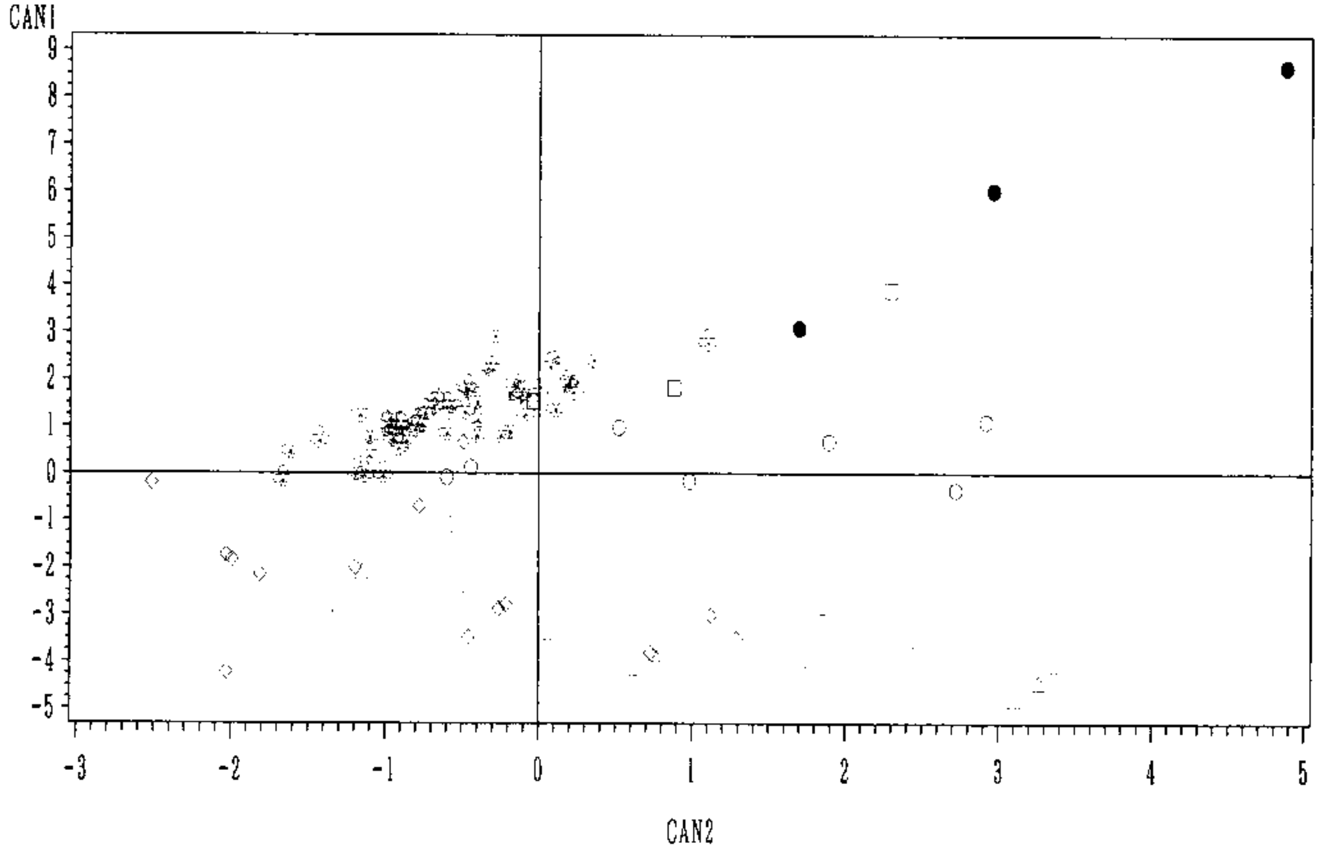

Fig. 4. Distribution of canonical scores of patients in model 3, in which the normal group was added to discriminant groups (AA, MDS, and so on) in model 2. Symbols used are the same as in Figure 3. Clover stands for control subjects.

indispensable for the further application of this analysis to other diseases. This is probably why only RBC was chosen in the discriminant analysis. Including WBC, two sets of laboratory results were constructed and applied for the further analyses: 1) RBC, MCH, Hb, PLT, and WBC, and 2) RBC, MCV, Ht, PLT, and WBC.

Estimation of Mahalanobis' generalized distances between these five disorders and normal controls revealed a rather close relationship between AA, MDS, and ITP. These results are considered reasonable, as MDS is located between anemia and leukemia, and further differentiation between AA and MDS is difficult without special morphological standards or chromosomal analyses [19]. Some ITP is difficult to differentiate from AA or MDS without special examinations, including bone-marrow analysis. Despite this, the three-dimensional analysis (Fig. 5) made it possible to differentiate between these quite well. In this series, we analyzed five disorders, in each of which at least 3 cases were available. However, there are other diseases, such as paroxysmal nocturnal hemoglobulinuria or myeloproliferative disorders, which are also close to MDS. Although the numbers of cases examined were too small and were not included in this series, the present analysis should be able to discriminate them from MDS or AA (data not shown).

On the other hand, PV and IDA were shown to be different from other disorders or controls and were almost always predicted correctly. One patient with PV was mistaken as having ITP, but this was a particular case of pseudothrombocythemia. In this study, we analyzed data from only 48 patients with five disorders, and 51 controls. The number of subjects and diseases may arguably have been too small, but the final canonical analysis, including patient properties, provided a quite satisfactory discrimination among diseases and controls. We should extend our studies further to increase the number of patients and diseases. New patients with these five disorders should be monitored with regard to this analysis, and other diseases should be checked for possible overlap with the five disorders. The latter should include considerable numbers of patients and diseases with secondary anemia. When certain overlaps are seen, we should consider other parameters of routine laboratory results, such as biochemical analyses, to discriminate them clearly from the patients with primary disorders of the hematopoietic system. Results of the present study and future studies should facili-

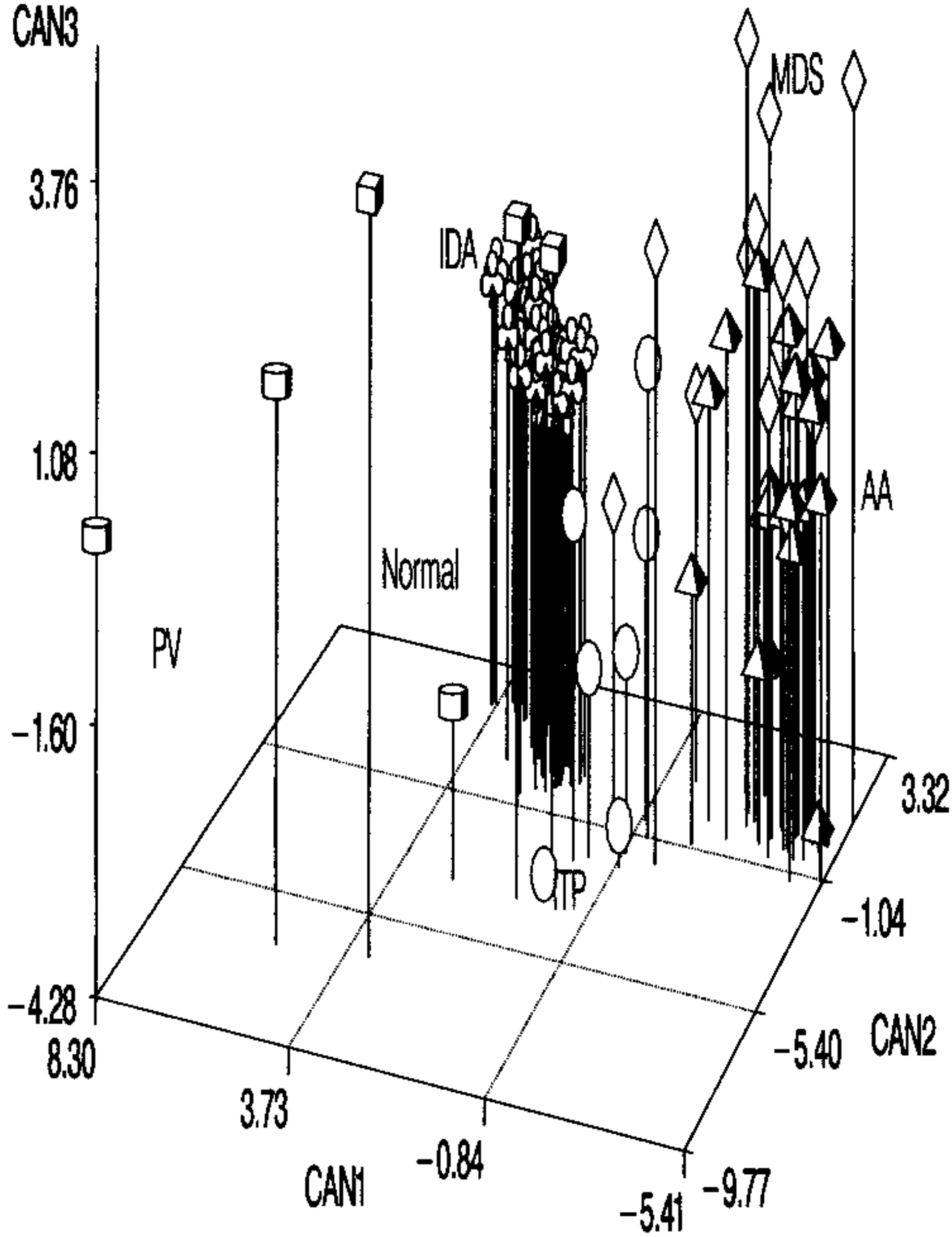

Fig. 5. Distribution of canonial scores of patients in model 4. Symbols used are the same as in Figure 4, except for PV, shown by cylinder.

tate production of a diagnostic compact disc, easily applicable on a personal computer at the sites of primary care, for general usage.

ACKNOWLEDGMENTS

The authors thank Mrs. K. Okamura and the staff of the Hematology Section of the Central Clinical Laboratory, Kyoto University Hospital, Kyoto, Japan for their technical assistance and valuable advice, and Miss Y. Mitsuda for secretarial assistance. This study was supported in part by a grant-in-aid from SRL, Inc. to Toshihito Furukawa. We also thank Mr. D. Mrozek for editing the manuscript in English.

REFERENCES

1. Matsuda N: Recent progress in computer assisted laboratory diagnosis for peripheral hemogram. Rinsho Byori [Suppl] 84:159, 1989.
2. Spiegelhalter DJ: Bayesian analysis in expert systems. Presented at the Sixtieth Annual Meeting of the Japanese Statistical Society, July 1992.
3. Houwen B: The use of inference strategies in the differential diagnosis of microcytic anemia. Blood Cells 15:509, 1989.

4. O'Connor ML, McKinney T: The diagnosis of microcytic anemia by a rule-based expert system using VP-expert. Arch Pathol Lab Med 113:985, 1989.
5. Pearson K: On lines and planes of closest fit to systems of points in space. Philosophical Magazine 6:559, 1901.
6. Hotelling H: Analysis of a complex of statistical variables into principal components. J Educ Psychol 24:417, 1933.
7. Kshirsager AM: Multivariate analysis. In Mardia KV, Kent JT, Bibby JM (eds): New York and London: Academic Press, 1979.
8. Kshirsagar AM: "Multivariate Analysis." New York and London: Marcel Dekker, Inc., 1979.
9. Lawley DN: Tests of significance in canonical analysis. Biometrica 46:5966, 1959.
10. Fisher RA: The use of multiple measurements in taxonomic problems. Ann Eugen 7:179, 1936.
11. Rao CR: "Linear Statistical Inference." New York: John Wiley & Sons, Inc., 1973.
12. Rao CR: "Linear Statistical Inference and Its Applications." New York: John Wiley & Sons, Inc., 1965.
13. Costanza MC, Afifi AA: Comparison of stopping rules in forward stepwise discriminant analysis. J Am Stat Assoc 74:777, 1979.
14. Jennrich RI: Stepwise discriminant analysis. In Enslein K, Ralston A, Wilf H (eds): "Statistical Methods for Digital Computers." New York: John Wiley & Sons, Inc., 1977.
15. Klecka WR: Discriminant analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-019. Beverly Hills: Sage Publications, 1980.
16. Rencher AC, Larson SF: Bias in Wilks' lambda in stepwise discriminant analysis. Technometrics 22:349, 1980.
17. Shiga S, Koyanagi I, Kannagi R: Clinical reference values for laboratory hematology tests calculated using the iterative truncation method with correction. Rinsho Byori 38:93, 1990.
18. Ichihara K, Miyai K: Graphical methods to analyse overall pathophysiological data structure from laboratory database. Rinsho Byori [Suppl] 83:39, 1989.
19. Werner M, Maschek H, Kalovtsi V, Choritz H, Georgii A: Chromosome analyses in patients with myelodysplastic syndromes: Correlation with bone marrow histopathology and prognostic significance. Virchows Arch [A] 421:47, 1992.